

**From:** [Goolsby, Rebecca CIV ONR, 341](#)  
**To:** ["Troyer, Lisa L CIV USARMY RDECOM ARL \(US\)"](#)  
**Cc:** [Bowman, Elizabeth K CIV \(US\) \(elizabeth.k.bowman.civ@mail.mil\)](#); ["Nitin Agarwal"](#); [huan.liu@asu.edu](#); ["kathleen.carley@cs.cmu.edu"](#); ["Justin Sampson"](#)  
**Subject:** RE: STTR - social media testbed/training data  
**Date:** Friday, February 10, 2017 11:48:00 AM

---

You don't ask for much, do you? :D

I want to think on this. Specifically, the use of two types of data that might provide some good instantiations of this. We're working on Blogtrackers and on finding the key "carriers" of anti-NATO discourse in Poland and the Baltics. This is a work in progress. Right now we're trying to identify key influential blogs that carry specific narratives --which would be of use to this effort but SO NOT READY yet. We're also examining Polish and Baltic Twitter (also in earliest stages). So if you needed this in six months, then we might have some stuff. Right now? Not so good. After april 1? Maybe, but it would be risky. This is a bit early for data sharing.

Second insight.

There's tons of English data, but that isn't the problem. The problem is Arabic, Russian and in general, languages other than English.

IF I WERE YOU:

I would constrain it to English and perhaps one other language for whom you can get outside expertise with appropriate social science credentials to help you. English and Russian for pro and anti-NATO discourses during Trident Juncture might be do-able, but you'd have to work with the performers (cc'd) --and I have overtasked these people already.

ASU has already packaged some Ukrainian / Crimea data from the invasion period in 2014 that might be of use and easily gotten. Twitter only.

Its social media so unvalidated, poorly geo-referenced, and full of fakery. But it is also full of different factions fighting an information during a period of relatively unsophisticated techniques.

We probably have Syria-related data, fairly unanalyzed, so no idea what's in there. Nitin has done a good bit of work on ISIS (and so has Kathleen).

That's all I got.

VR, RG

-----Original Message-----

From: Troyer, Lisa L CIV USARMY RDECOM ARL (US) [<mailto:lisa.l.troyer.civ@mail.mil>]  
Sent: Friday, February 10, 2017 9:15 AM  
To: Bowman, Elizabeth K CIV USARMY RDECOM ARL (US); Goolsby, Rebecca CIV ONR, 341  
Subject: STTR - social media testbed/training data

Hi Liz, Rebecca,

I hope you're both doing well. I'm working with Tim Perkins from ERDC on a STTR, "Beliefs, Values, Interests and Practices of Identities, Networks, Groups for Planning & Analysis," and we are moving toward Phase II proposals, following the Phase I close-out.

I'm writing to ask you if you can point me to any existing DoD-developed/sponsored testbed or training data for social media analytics that you're aware of. If it is possible, we would like to see if our Phase II teams can gain access to such testbed data to facilitate the development of their tools/systems/analytics. Ideally, the data would be cross-cultural, validated, and geocoded, although if not all of these requirements are met, then it may still be of value to the teams.

In Phase II, the performers will be asked (among other things) to validate their systems for analyzing data to generate interpretations of identities, beliefs, values in social networks. While they are not relying exclusively on social media, it does play a role for all of the current teams that will compete for a Phase II award.

Thanks for any expert guidance you can provide! I'm happy to tell you more about the teams (both ONR and ARL were represented at the Phase I close-out meeting presentations).

Have a pleasant weekend.

Sincerely,

Lisa

Lisa Troyer, Ph.D.  
Program Manager  
Social & Behavioral Sciences Program  
Life Sciences Division  
Army Research Office  
Office: (919) 549-4230  
Cell: (919) 628-8061  
lisa.l.troyer.civ@mail.mil